Cancer Imaging (2000) **1,** 28–34

Keynote Lectures — Importance of Tumour Measurements

Tuesday 10 October 2000, 09.15–10.45

Cancer imaging — the significance of the findings

Rodney H Reznek

Academic Department of Radiology, St Bartholomew's Hospital, London, UK

Introduction

Over the past several years, the technological advances in the field of body imaging have been almost too numerous to catalogue. Each modality goes through a cyclical pattern of evolution. In the earliest phase of this evolution, most research is descriptive and anecdotal in nature. As a modality becomes established, it enters the second phase, in which it is touted as being superior to all prior conventional modalities. The third phase represents a backlash effect in which the shortcomings of a new technique and its inferiority to prior modalities are stressed. Finally, the technical development of a new modality reaches a plateau, the literature reflects an equality with earlier techniques. It is during this phase that the true cost-effectiveness of a new imaging technique and its impact on patient care are established. All new imaging modalities go through these phases. Most academic radiology departments are composed of single modality advocates who fail to see the interrelationships among the available imaging techniques. The radiologists must be prepared to offer unbiased aid to the referring clinicians in choosing the most cost-effective procedure from the radiological armamentarium. Nowhere is this more relevant than in imaging patients who have cancer where the range of anatomy and pathology to be imaged is infinite. Close co-operation between the clinician and radiologist is essential and a clear understanding of the purpose of the imaging is mandatory.

The aims of imaging

Diagnostic imaging fulfils several functions in patients with cancer. These include making a diagnosis; staging the disease; monitoring response; detecting recurrence and research applications.

Diagnosis

It is only infrequently that straightforward imaging provides a sufficiently specific diagnosis on which treatment can be based. However, the application of image-guided biopsy techniques has revolutionized the ease with which cytology or histology can be obtained. Few anatomic sites are now inaccessible to the skilled radiologist using imaging guidance. The choice of the most appropriate form of imaging guidance will vary from institution to institution depending on the skill and preference of individual radiologists and also on the site of the disease.

Staging the disease

Increasingly, imaging techniques are being applied to provide a more refined and accurate staging of the disease. However, to do this requires a detailed knowledge of the sensitivities, specificities and accuracy of individual imaging techniques as they relate to assessing the stage of each individual pathological process. These will be discussed in detail below. A detailed knowledge of the advantages and limitations of each imaging technology as they apply to the assessment of each individual stage is required. In order to do this, close collaboration between clinician and radiologist is often essential. For example, the simple assessment of the possibility of liver metastases requires a knowledge of the most appropriate technique, its accuracies, pitfalls and shortcomings in assessing focal pathology.

Monitoring response and detecting recurrence

Increasingly, with the more effective use of radiotherapy and chemotherapeutic agents, imaging is used to assess the response to therapy and to detect recurrent disease. This usually relies on a straightforward assessment of a change in size although occasionally a change in the characteristics of a mass lesion can also provide information as to a changing response. It is recognized, however, that this is a crude form of assessment to response and that imaging is but one facet in the overall

1470-7330/00/010028+07

assessment of the patient's response to therapy. Frequently, changes in the imaging appearance result from the effect of radiotherapy or chemotherapy and a detailed knowledge of these appearances is required both by the radiologist and by the oncologist. Similarly, an appreciation of the phenomenon of a residual 'sterile' mass is also necessary together with possible imaging strategies for evaluating this residual mass.

Research applications

The very high accuracy and reproducibility of many imaging techniques make it extremely well suited to Phase II trials in which the oncologist is assessing the biological activity of new treatments. In Phase III trials, when comparing the results of different treatments, survival is usually the final arbiter. If the size of the patient group is large enough, sophisticated staging is not needed as the stage will be randomized out. But in practice, the groups tend to be small and one of the prognostic variables, namely, the varying stage of the diseases can be removed from the study by achieving more accurate staging through CT. Another application in Phase III trial is in advanced disease where there is no obvious difference in survival and one is looking not for survival but an increase in biological activity. So, in this sort of study of patients with advanced disease, where the end point of the study is a response rate, not survival, imaging becomes a very valuable tool because of its accuracy in monitoring the disease.

Choice of the appropriate technique

The choice of the most appropriate technique for each particular aim in assessing patients with cancer, depends on several factors: these include an assessment of the sensitivity and specificity, the availability of any technique, the cost and the cost-effectiveness.

Sensitivity and specificity

When selecting a diagnostic test, one of the most important considerations is the accuracy of the test. Diagnostic accuracy is best described in terms of sensitivity and specificity. Stated simply, sensitivity is the ability of the test to recognize disease and specificity is the ability of the test to recognize normality. These concepts are both illustrated with the help of a Binary Table that depicts the correlation between test interpretation and the presence or absence of disease in the population under study.

The Binary Table categorizes patients into four mutually exclusive outcomes:

(1) positive test result in the disease present, true positive (TP);
(2) positive test result in the disease absent, false positive (FP);
(3) negative test result in the disease present, false negative (FN); and
(4) negative test result in the disease absent, true negative (TN).

The sensitivity of the test is the proportion of patients with disease who have a positive test result. In other words, it is the ability of the test to recognize disease.

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP}+\text{FN}}$$

The specificity of a test is the proportion of patients without disease who have a negative test result. In other words, it is the ability of the test to recognize normality or the absence of a particular disease.

$$\text{Specificity}=\frac{\text{TN}}{\text{TN}=\text{FP}}$$

The accuracy of the test is of less value than the sensitivity and specificity because it lumps together positive and negative results.

The positive predictive value of a test indicates the probability of whether the disease is actually present if the test is positive.

$$\text{Positive predictive value (PPV)}=\frac{\text{TP}}{\text{TP}+\text{FP}}$$

$$\text{Negative predictive value (NPV)}=\frac{\text{TN}}{\text{TN}+\text{FN}}$$

Receiver–Operator–Characteristics analysis

The sensitivity and specificity of a test depend on the criteria chosen for interpretation. As the criteria for calling a test result positive are made more stringent, specificity improves at the expense of sensitivity. Conversely, as the criteria are relaxed, sensitivity improves while specificity diminishes. This relationship can be demonstrated on a receiver–operator–characteristics (ROC) curve. This curve is generated by plotting the sensitivity (true-positive rate) this is 1 — for

		DISEASE	
		Present	Absent
TEST	Positive	True-positive	False-positive
	Negative	False-negative	True-negative

***Table 1** Interobserver basic data for assessment of 85 xeromammograms by two radiologists, after Altman and taken from a larger study by Boyd et al.*

Radiologist A	Radiologist B				
	Normal	Benign	Suspected cancer	Cancer	Total
Normal	21	12	1	0	33
Benign	4	17	1	0	22
Suspected cancer	3	9	15	2	29
Cancer	0	0	0	1	1
Total	28	38	16	2	85

specificity (false-positive rate) for the different interpretation criteria. The fundamental principle illustrated by the ROC curve is that there is an inherent limit to the diagnostic accuracy of a test. Once this limit has been reached, the interpreter can only improve sensitivity at the expense of specificity and vice versa. The ROC curve can be used to select the 'best' cut-off criteria for positivity, taking the pre-test probability and the relative cost (in terms of patient outcome) of false-positive and false-negative test results into account. Additionally, ROC curves are useful in comparing the performance of different tests, because they allow for a wide range of different positivity (criteria).

Interobserver agreement (kappa test)

Altman (see further reading) describes well how to measure interobserver agreement, using as data the assessments of 85 xeromammograms by two radiologists (A and B) where the xeromammogram reports are given as one of four results: normal; benign disease; suspected cancer; cancer.

A measure of agreement is required between radiologist A and radiologist B rather than a test of association such as might be undertaken using the χ^2 test.

As Altman points out, the simplest approach is to count how many exact agreements were observed between A and B, which from Table 1 is 54/85=0.64. However, the disadvantages with this method of merely quoting a 64% measure of agreement are that it does not take into account where the agreements occurred and also the fact that one would expect a certain amount of agreement between radiologist A and radiologist B purely by chance, even if they were guessing their assessments.

The expected frequencies along the diagonal of Table 1 are given in Table 2 from which it is seen for these data that the number of agreements expected by chance is 26.2 which is 31% of the total, i.e. 26.2/85. What the kappa test gives is the answer to the question of how much better the radiologists were than 0.31.

The maximum agreement is 1.00 and the kappa statistic gives the radiologists' agreement as a proportion of the possible scope for performing better than chance, which is $1.00 - 0.31$.

$$\kappa = (0.64 - 0.31)/(1.00 - 0.31) = 0.47$$

There are no absolute definitions for interpreting κ but it has been suggested that the guidelines in Table 3 can be followed, which in the example considered here means that there was moderate agreement between radiologist A and radiologist B.

Stage migration ('Will Rogers' effect)

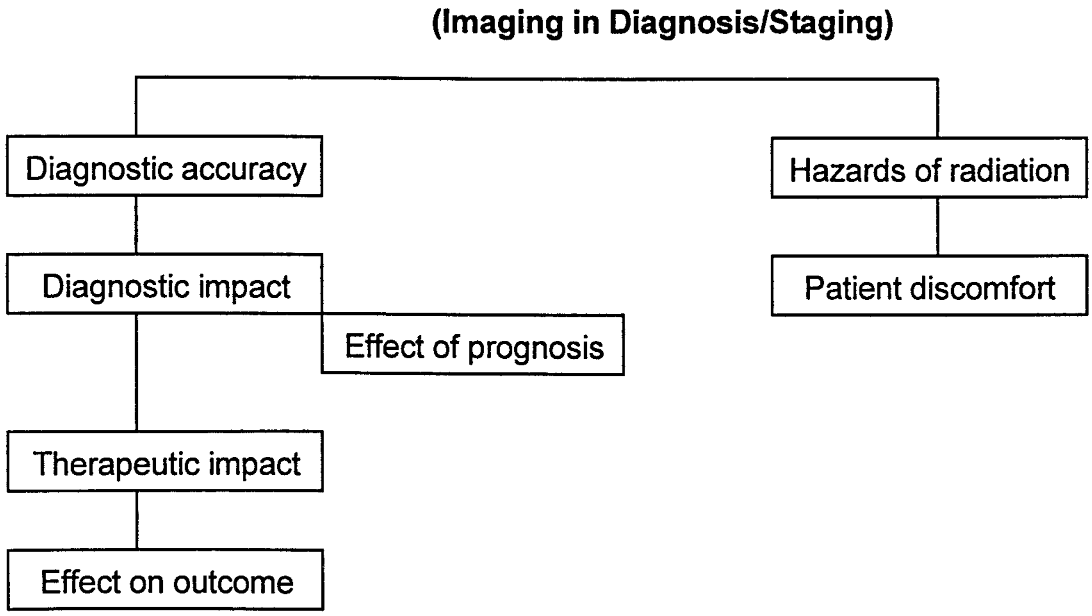
An important impact of the use of sophisticated techniques to stage patients with cancer is the apparent continuous improvement in cancer survival rates reported over the last 25 years. Although this is quickly and easily attributable to earlier diagnosis and new and more effective treatments, the effect of more accurate staging may to some extent explain these improved results. Feinstein *et al*. found that a 1977 cohort of patients who had undergone lung cancer treatment survived significantly longer in each of three TNM subcategories than a cohort managed in the 1950s and 1960s; a finding which is not surprising. When, however,

***Table 2** Calculation of the expected frequencies for the kappa test, after Altman*

Assessment	Expected frequency
Normal	$33 \times (28/85) = 10.87$
Benign	$22 \times (38/85) = 9.84$
Suspected cancer	$29 \times (16/85) = 5.46$
Cancer	$1 \times (3/85) = 0.04$

***Table 3** Guidelines for the interpretation of the κ statistic (REF)*

κ values	Strength of agreement
<20	Poor
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Very good



he staged the recent cohort on clinical grounds only — without the benefit of ultrasonography, CT and nuclear medicine — these survival differences disappeared. It was apparent that the improved survival rates were mainly an artefact of better staging; patients in the lower stages with clinically occult (usually nodal) disease were being identified with better imaging and were being placed in a more advanced stage ('stage migration'). Better staging leads to benefit to all; in the lower stages, patients with occult metastases would be removed with benefit to those stages; in the higher stages, those patients with a lower tumour burden would be added to those with a higher burden, with improvement in survival rates. Thus while individual prognosis did not change overall, survival in each stage improved. The stage migration phenomenon occurs when comparisons are made between groups of patients who have undergone less or more thorough staging techniques and as such is likely to occur when the comparisons are made over a time period which spans the introduction of new technology. It has been noted with numerous tumours including metastatic germ cell tumours and gastric cancers.

Diagnostic procedure

The diagnostic accuracy of most techniques is high but irrespective, figures for accuracy are readily available. The diagnostic impact is not limited to a change in diagnosis or prognosis but includes the ease with which the diagnosis is reached, reduction in the number of invasive investigations, and reduction of the time spent in hospital. It is self-evident that by achieving percutaneous biopsies, and diagnosing and staging tumours accurately without numerous more invasive investigations (including surgery) that most forms of imaging can be of benefit to the patient. As regards therapeutic impact there are several studies showing that imaging substantially alters the patient's management. The effect on patient outcome is a great deal more difficult to evaluate than any of these other factors. In the short term, it has a very obvious impact by reducing the number of invasive test, by reducing the time in hospital and by avoiding surgery. The long-term effect, such as an improvement in the rate of cure, or the rate of survival or even the relief of symptoms brought about by the imaging technique is a great deal more difficult to evaluate.

Further reading

Altieri V, Setola P, Ottaviano N *et al.* Imaging diagnosis of non-lymph node metastasis of bladder carcinoma. Arch Ital Urol Androl 1994; 66: 223–8.

Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991; 404–9.

Armstrong P, Black WC. Optimum utilisation of radiological tests: the radiologist as advisor. Clin Radiol 1989; 40: 444–7.

Basseres N, Grob JJ, Richard MA *et al.* Cost-effectiveness of surveillance of Stage 1 melanoma. A retrospective appraisal based on a 10-year experience in a dermatology department in France. 12th Annual International Breast Cancer Conference, March 1995, Miami, FL, 1995.

Biggs CG, Ballantyne GH. Sensitivity versus cost effectiveness in postoperative follow-up for colorectal cancer. Curr Opin Gen Surg 1994; 94–102.

Black WC, Armstrong P, Daniel TM. Cost effectiveness of chest CT in T1N0M0 lung cancer. Radiology 1988; 167: 373–378.

Bragg DG. The impact of imaging technology on cancer survival 1970 to 1992. Invest Radiol 1993; 28: S132–S133.

Bosi GJ, Geller NL, Chan EY. Stage migration and the increasing proportion of complete responders in patients with advanced germ cell tumours. Cancer Res 1988; 48: 3524–7.

Boyd NF, Wolfson C, Moskowitz M. Observer variation in the interpretation of xeromammograms. J Natl Cancer Inst 1982; 68: 357–63.

Bunt AMG, Hermans J, Smith VTHBM *et al*. Surgical/pathologic stage migration confounds comparisons of gastric cancer survival rates between Japan and Western countries. J Clin Oncol 1955; 13: 19–25.

Colice GL, Birkmeyer JD, Black WC *et al*. Cost-effectiveness of head CT in patients with lung cancer without clinical evidence of metastases. Chest 1995; 108: 1264–71.

Dixon AK, Southern JP, Teale A *et al*. Magnetic resonance imaging for the head and spine: effective for the clinician or the patient? Br Med J 1991; 302: 78–82.

Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. N Engl J Med 1985; 312: 1604–8.

Fineberg HV, Wittenberg J, Ferrucci JT. The clinical value of body computed tomography over time and technologic change. Am J Roentgenol 1983; 141: 1067–72.

Goldin J, Sayre JW. Review: a guide to clinical epidemiology for radiologists: part ii statistical analysis. Clin Radiol 1996; 51: 317–24.

Kahn CE, Sanders GD, Lyons EA *et al*. Computed tomography for nontraumatic headache: current utilization and cost-effectiveness. Can Assoc Radiol J 1993; 44: 189–93.

Kairaluoma MI, M Valtteri, Partio E *et al*. Impact of new imaging techniques on survival in cancer of the head of the pancreas and the periampullary region. Acta Chir Scand 1985; 151: 69–72.

Kelsey Fry I. Who needs high technology? Br J Radiol 1984; 57: 765–72.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–74.

Larson SM. Improving the balance between treatment and diagnosis: a role for radioimmunodetection. Cancer Res 1995; 55(23 Suppl): 5756s–5758s.

MacKenzie R, Dixon AK. Measuring the effects of imaging: an evaluative framework. Clin Radiol 1995; 50: 513–8.

Maroldi R, Farina D, Battaglia G *et al*. Magnetic resonance and computed tomography compared in the staging of rhinosinusal neoplasms. A cost-effectiveness evaluation. Radiol Med (Torino) 1996; 91: 211–8.

Robson AK, Leighton Sem Anslow P, Milford CA. MRI as a single screening procedure for acoustic neuroma: a cost effective protocol. J Roy Soc Med 1993; 86: 455–7.

Wiggers T, Wagenaar H, de Charro FT. CEA directed or symptomatic follow-up in colorectal cancer. A theoretical cost-effectiveness analysis. Rev Oncol 1993; 3: 39.

International criteria for tumour assessment

L Ollivier

Institut Curie, Paris, France

Introduction

The number of patients treated for cancer is increasing regularly. This is partially due to an earlier detection of the disease and improved survival linked to new cytotoxic agents. Because of the cost and toxicity of these treatments, a rigorous and accurate evaluation of their efficacy is necessary, as well as the evaluation of the toxicity which is already carried out.

Most patients treated for malignancies are treated using protocols with drugs whose efficacy and toxicity have already been established. Diagnostic imaging plays a major role in the follow-up of these patients, but the progression of disease or recurrence are assessed radiologically and clinically.

Patients with advanced disease may be treated with new cytotoxic agents evaluated in clinical trials. In these trials, the tumour shrinkage or tumour response based on the decrease in size of the lesions is still a common end-point.

The need for rigorous assessment of drug efficacy has been emphasized in the literature from the early days of chemotherapy, when it became necessary to compare the results of different teams obtained in clinical trials. In these cases, diagnostic imaging supports clinical research and the imaging modalities have to be compared by an evaluation committee. Guidelines and international rules on response evaluation were progressively established during the 1970s. The WHO (World Health Organization) guidelines[1] are the most widely used. These guidelines and rules have been written by oncologists and statisticians, but radiologists were not involved in their elaboration. Now these rules have become obsolete and need to be updated in the light of improvements in the accuracy of imaging tools.

Definitions

Overall survival and objective response rates are the usual parameters used to assess response to treatment during clinical trials in oncology. Overall survival is the gold standard in oncology but the delay necessary to obtain this parameter is sometimes too long: the physicians need to determine rapidly whether the agent demonstrates encouraging results or not, in order to adjust the therapy or to include the patient in another trial. But, in contrast to survival, objective response (or tumour shrinkage, or tumour response) is more difficult to assess because it is highly dependent on the quality of clinical and radiological tumour measurements.

Recommendations have been reported in the literature to measure tumoral masses[1–3], also called targets, but many factors interfere with response evaluation, such as the quality and reproducibility of the radiological examinations, the choice of targets and the investigator's objectivity[3–5].

In the WHO's criteria, tumour lesions will be considered as measurable disease (bidimensionally or unidimensionally measurable, all the measurements being recorded in metric notation) and evaluable or unmeasurable disease (i.e. lymphangitis, skin involvement, abdominal masses palpated but not measured, etc.).

To evaluate the tumour response it is essential to follow a reliable and reproducible method of measurement. Except for spherical lesions, it is not possible to measure the volume or even the surface of a target precisely. The surface area approximation of the target is measured by multiplying the longest diameter by the greatest perpendicular diameter.

The tumoral volume is defined by the international guidelines as the sum of the surfaces of the targets. It is in fact the sum of an approximation of tumoral surfaces and seems to be an arbitrary definition, but it makes it possible to compare the evolution of the tumoral volume and to compare the results obtained by different teams.

Definitions of objective response in measurable disease

Complete response (CR): the disappearance of all known disease, determined by two observations not less than 4 weeks apart. In addition there can be no appearance of new lesions.

Partial response (PR): 50% or more decrease in total tumour size of the lesions which have been measured to determine the effect of therapy by two observations not less than 4 weeks apart. No appearance of new lesions or progression of any lesion.

Stable disease (SD): A 50% decrease in total tumour size cannot be established nor has a 25% increase in the size of one or more measurable disease lesions been demonstrated. No appearance of new lesions.

Progressive disease (PD): A 25% or more increase in the size of one or more measurable lesions, or the appearance of new lesions.

It is clear that the complete response is easy to assess, but that the rate of 50% is arbitrary. It is sometimes difficult to be so accurate, particularly when the lesion is single, of very small size, or when the lesions are multiple, or are infiltrative. In the WHO's guidelines, there are no recommendations about the minimal size of the lesion, or the number of lesions to consider. When lesions are measured with electronic calipers directly at the console, the accuracy of the measurements is better, but in the late 1970s, cross-sectional imaging was not so widely used and lesions were measured on plain film with a ruler. A separate set of response criteria have been defined for bone metastases but it is so complicated that it is practically never used. Bone metastases are now excluded from most reliable trials.

Reasons for inter-observer (and intra-observer) variability

Recently, a French group studied the impact of an evaluation committee (EC) on patients' overall response status in a large multicentre trial in oncology[5]. They identified reasons for disagreements between investigators and the EC. Overall tumour response was reduced by 23.2% after the review by the EC. Reasons for major disagreements included errors in tumour measurements and errors in selection of measurable targets. Pitfalls such as tumoral necrosis, intercurrent diseases, and radiological technical problems were discussed. They concluded that all therapeutic trial results should be reviewed by an independent EC[5,6].

On the basis of this experience the Group of Clinical Evaluation (GREC) was created in France in 1996; this group is working with the EORTC (European Organization for Research and Treatment of Cancer) on the elaboration of new guidelines[7].

The imaging modalities

All imaging strategies can be used in oncology, particularly cross-sectional imaging, but in daily practice, conventional imaging and clinical findings may be sufficient to assess the evolution of illness; in contrast, in clinical trials, cross-sectional imaging has to be used with a consistent technique throughout the trial, and ultrasound is not accepted in most of the trials because of its operator dependence. The use of MRI is sometimes avoided because this rather new technique is not widely performed. Precise tumour evaluation in a therapeutic setting may require a different procedure (slice thickness, contrast media, etc.) from that used for standard tumour diagnosis or routine tumour evaluation for an individual patient.

Evolution

Axial measurements fail to recognize significant variations during treatment of tubular-like lesions, especially when mainly the third (cranio-caudal) tumour diameter is altered. Cross-section imaging readily allows the calculation of the three main axis perpendicular diameters. Several therapeutic evaluation protocols already use the three maximal diameters which are considered to analyse the tumour course more accurately. On the other hand, members of the EORTC and of the NCI (National Cancer Institute) Canada, recently published a controversial study[8] concluding that unidimensional measurement of tumour maximum diameter might be sufficient to assess change in solid tumours. Moreover, it is less laborious and the risk of errors is decreased. On the basis of this demonstration, the authors then reviewed the WHO criteria and proposed new guidelines to evaluate the response to treatment.

Other considerations than the tumoral volume may lead to other end-points[9] such as surrogate markers, quality of life, and also the composition of the tumour, especially the residual masses. MRI or PET-scans are promising techniques, but not widely used and not recommended in clinical trials. As long as the response rate is still an end-point in therapeutic trials, measurements of the tumoral volume are mandatory and necessitate updated guidelines.

Revisited version of the WHO's criteria of response in solid tumours

After several years of intensive discussions, a new set of guidelines has been defined in a special article published in the *Journal of the National Cancer Institute* and previously presented to the scientific community at the ASCO (American Society for Clinical Oncology). These criteria support the simplification of response evaluation through the use of unidimensional measurements and the sum of the longest diameters instead of the conventional method using two measurements and the sum of the products. The other guidelines introduce the use of computed tomography (CT) and magnetic resonance imaging (MRI): technical recommendations are provided in the article concerning the slice thickness, the use of contrast media, etc. Ultrasound should not be used to measure tumour lesions or as a possible alternative to clinical measurements for superficial palpable lesions. The new response criteria are linked to the relationship between change in diameter, product and volume: the response which was defined as a 50% decrease using the previous WHO criteria becomes 30% with the new criteria (diameter) and the disease progression becomes 20% instead of 25%. This correlation has been done in order to allow comparison with the response rates obtained with the WHO criteria, particularly in historical trials.

Evaluating clinical research cost

Participation in clinical research is not only time-consuming for the department of diagnostic imaging, it also represents a huge amount of money. Experimental protocols demand more than standard treatments. But how much more? This question has been answered by the clinical research unit of a French oncology department[10]. They compared the costs of two clinical trials with standard treatment. Not only the price of the drugs were compared but also the costs incurred by the extra time taken by the oncologists to inform and to enrol the patient, and by the nurses to adjust the new therapies. The assessment costs were also evaluated (laboratory tests and radiological evaluation). The extra time used by radiologists to perform the mandatory examinations (depending on the frequency of tumour re-evaluation), to measure the lesions precisely, and to compare the successive examinations was taken into account. The price of supplementary films or contrast media, and of the optical disks used to record the examinations, was calculated. This was an important study for the radiologists: they could then ask for greater financial support, and more staff in their department.

Conclusion

Evaluation of the efficacy of anti-tumour treatments with modern medical imaging is becoming a more precise and more complex activity that requires collaboration on the part of radiologists and clinicians who should be trained in these techniques, as the objectives and methods differ from those of standard diagnosis and management. New guidelines have been proposed recently, as the result of large-scale international collaboration. They have to be validated by the scientific community and amended by further trials. Each therapeutic trial protocol should be initially validated by experts in evaluation and an independent review committee should review the patients' files and radiological images of at least the presumed responders.

References

[1] World Health Organization. WHO Handbook for Reporting Results of Cancer Treatment. Geneva: World Health Organization, 1979: 48.

[2] Ollivier L, Thiesse P, Leclère J *et al*. Rôle de l'imagerie dans l'évaluation de la réponse sous chimiothérapie. La Lettre du Cancérologue 1999; VIII: no. 1.

[3] Husband JE. Monitoring tumour response. Eur Radiol 1996; 6: 775–83.

[4] MacVicar D, Husband JE. Assessment of response following treatment for malignant disease. Br J Radiol 1997; 70: 41–7.

[5] Thiesse P, Ollivier L, Di Stefano-Louineau D *et al*. Response rates accuracy in oncology trials: reasons for inter-observer variability. J Clin Oncol 1997; 15: 3507–14.

[6] Escudier B, Ollivier L, Thiesse P *et al*. Why should response rates be controlled by an evaluation committee. Proc ASCO 1996; 15: abstr.316.

[7] JamesK, Eisenhauer E, Christian M *et al*. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999; 91: 523–8.

[8] Therasse P, Arbuck S, Eisenhauer E *et al*. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000; 92: 205–16.

[9] Gelmon KA. The fine points of endpoints: phase II trials in lung cancer. Ann Oncol 1998; 9: 1045–6.

[10] Bercez-Leflour C, Viens G, Bonneterre ME, Bonneterre J. Evaluating clinical research costs. Applied Clinical Trials 1996; 11: 54–8.